

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 1, 2003, 08:48:11 ; Search time 22.5 Seconds

(without alignments)
5120.944 Million cell updates/sec

Title: US-09-768-781-2

Perfect score: 2543

Sequence: 1 atgaacacaagaccacaaca.....caaggcaaaagtgtgtctga 1389

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09768781/runat.01042003.084729.4929/app.query.fasta_1.1543
-DB=SwissProt40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPEGL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09768781 @CGN 1 1 31 @runat.01042003.084729.4929 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	939	36.9	444	1 XKY_HUMAN	P51811 homo sapien
2	197.5	7.8	159	1 XKY_HUMAN	O14609 homo sapien
3	108.5	4.3	497	1 DTPT_LACCHI	O07380 lactobacill
4	108.5	4.3	741	1 YBTO_ECOLI	P75783 escherichia
5	104.5	4.1	521	1 Y725_CABEL	Q10934 caenorhabdi
6	102	4.0	1584	1 BAIT1_HUMAN	O14514 homo sapien
7	100.5	4.0	264	1 ATP6_PODAN	P15994 podospora a
8	100.5	4.0	459	1 NUAM_CERSI	O03204 ceratotheri
9	99.5	3.9	551	1 YNBI_YEAST	P53730 saccharomyc
10	99	3.9	233	1 ATP1_OCHNE	Q40607 ochrophaer
11	99	3.9	458	1 ACHO_HUMAN	Q05901 homo sapien
12	98	3.9	805	1 VPH1_SCHPO	O13742 schizosacch
13	97.5	3.8	325	1 MCSR_RAT	P35345 rattus norv
14	97	3.8	382	1 NUZM_CHLRE	P08740 chlamydomon
15	97	3.8	527	1 NUSM_CAEEL	P24896 caenorhabdi
16	96.5	3.8	384	1 NUSM_ANOAR	P51899 anopheles a
17	96.5	3.8	580	1 NUSM_ANOGA	P34854 anopheles g
18	96	3.8	371	1 NYGR_MOUSE	Q61212 mus musculu

19	96	3.8	1033	1 YD56_SCHPO	Q10309 schizosacch
20	95.5	3.8	252	1 ATP6_NEUCR	P37212 neurospora
21	95.5	3.8	255	1 ATP6_YARLI	Q36258 yarrowia li
22	95.5	3.8	325	1 MCSR_MOUSE	P41149 mus musculu
23	95.5	3.8	405	1 YPMO_YEAST	Q03687 saccharomyc
24	95.5	3.8	429	1 SECY_AQUAE	O66491 aquifex aeo
25	95	3.7	469	1 NUON_BUCAI	P57264 buchnera ap
26	94.5	3.7	438	1 MNT2_PSEAE	Q9762 pseudomonas
27	94.5	3.7	449	1 G64F_DROME	P83297 drosophila
28	94.5	3.7	468	1 CBIA_FUGRU	Q98894 fugu rubrip
29	94.5	3.7	786	1 APE1_SULTO	O96429 sulfolobus
30	94.5	3.7	865	1 CYAA_PROMI	O59685 proteus mir
31	94	3.7	592	1 FZD1_CHICK	O57328 gallus gall
32	94	3.7	607	1 NUSM_MOUSE	P03921 mus musculu
33	94	3.7	692	1 NUSC_WARPO	P06284 marchantia
34	94	3.7	715	1 LCCL_LACLA	Q96388 lactococcus
35	93.5	3.7	456	1 G64A_DROME	P83293 drosophila
36	93.5	3.7	971	1 CSE1_MOUSE	O9erk4 mus musculu
37	93.5	3.7	1564	1 PDRA_YEAST	P51533 saccharomyc
38	93	3.7	406	1 HOF2_HAEIN	P44821 haemophilus
39	93	3.7	497	1 NUZM_CHOCR	P48903 chondrus cr
40	93	3.7	871	1 YAYD_SCHPO	Q09766 schizosacch
41	92.5	3.6	327	1 GPS2_HUMAN	Q13227 homo sapien
42	92	3.6	248	1 ATP1_WARPO	P06289 marchantia
43	92	3.6	437	1 NUAM_ALBCO	P48914 albinaria c
44	92	3.6	455	1 ACHO_CHICK	P43679 gallus gall
45	92	3.6	554	1 NUSM_APLIL	P34855 apis mellif

ALIGNMENTS

RESULT 1

XX HUMAN
ID XX HUMAN STANDARD; PRT; 444 AA.
AC P51811;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Membrane transport protein XK (Kx antigen).
GN XK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94273191; PubMed=8004674;
RA Ho M., Chelly J., Carter N., Danek A., Crocker P., Monaco A.P.;
RT "Isolation of the gene for McLeod syndrome that encodes a novel
membrane transport protein.";
RL Cell 77:869-880(1994).
RN [2]
RP REVISIONS TO 204-205.
RA Ho M.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN SODIUM-DEPENDENT TRANSPORT OF NEUTRAL
AMINO ACIDS OR OLIGOPEPTIDES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: HIGH LEVELS IN SKELETAL MUSCLE, HEART, BRAIN,
AND PANCREAS; LOW LEVELS IN PLACENTA, LUNG, LIVER, AND KIDNEY.
CC -!- POLYMORPHISM: XK IS RESPONSIBLE FOR THE KX BLOOD GROUP SYSTEM.
CC -!- DISEASE: DEFECTS IN XK ARE THE CAUSE OF MCLEOD SYNDROME, AN X-
LINKED MULTISYSTEM DISORDER CHARACTERIZED BY ABNORMALITIES IN THE
NEUROMUSCULAR AND HEMATOPOIETIC SYSTEMS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z32684; CAA83632.2; --
 DR Genew; HGNC:12811; XK.
 DR MIM; 314850; --
 KW Transmembrane; Transport; Amino-acid transport; Blood group antigen.
 FT DOMAIN 1 2
 FT TRANSMEM 3 23 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 24 37 POTENTIAL.
 FT TRANSMEM 38 58 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 59 68 POTENTIAL.
 FT TRANSMEM 69 89 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 90 140 POTENTIAL.
 FT TRANSMEM 141 161 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 162 171 POTENTIAL.
 FT TRANSMEM 172 192 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 193 208 POTENTIAL.
 FT TRANSMEM 209 229 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 230 235 POTENTIAL.
 FT TRANSMEM 236 256 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 257 277 POTENTIAL.
 FT TRANSMEM 278 298 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 299 317 POTENTIAL.
 FT TRANSMEM 318 338 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 339 349 POTENTIAL.
 FT TRANSMEM 350 370 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 371 444 POTENTIAL.
 SQ SEQUENCE 444 AA; 50902 MW; 6F90B0B45659AIDA CRC64;

Alignment Scores:

Pred. No.: 1.16e-69 Length: 444
 Score: 939.00 Matches: 180
 Percent Similarity: 64.68% Conservative: 80
 Best Local Similarity: 44.78% Mismatches: 132
 Query Match: 36.92% Indels: 10
 DB: 1 Gaps: 4

US-09-768-781-2 (1-1389) x XK_HUMAN (1-444)

QY 136 TTTCATTATGACATCTTTCTCCACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTG 195
 Db 3 PheProAlaSerValLeuAlaSerValPheLeuPheValAlaGluThrThrAlaAlaLeu 22
 QY 196 TACATGTTAGATCTATCGAAGAATAGTGAACTTACTGTGATGACATACACCTTTTCT 255
 Db 23 SerLeuSerSerThrThrArgSerGlyGlyAspArgMetTrpGlnAlaLeuThrLeuLeu 42
 QY 256 TTCTTATGTTTTCATCCATTATGTCCTGACCTGACCTCTTTTGTCCACAGAGATCTA 315
 Db 43 PheSerLeuLeuProCysAlaLeuValGlnLeuThrLeuPheValHisArgAspLeu 62
 QY 316 GCCAAGATAAACGGCTATCATTTATGATCTAATCTCTCTGGGACCTGTTATCAGA 375
 Db 63 SerArgAspArgProLeuValLeuLeuHisLeuLeuGlnLeuGlyProLeuPheArg 82
 QY 376 TGTGTGGAGCCATGATTAAGTACTCTCACACTGTGGGAAGAAGAGGAGGAGGAGGCC 435
 Db 83 CysPheGluValPheCysIleTyr-----PheGlnSerGlyAsnAsnGluGluPro 99
 QY 436 TATGTGAGCTCAACCCGAAGAAG--ATGCTATATAGTCGCGAGGAGTCTCATAGAA 492
 Db 100 TyrValSerIleThrLysArgGlnMetProLysAsnGlyLeuSerGluGluLeuGlu 119
 QY 493 TGGAGGTGGGCCACTCCATCCGGACCTGGCTATGACCGCAATGCTCAACACGATG 552
 Db 120 LysGluValGlyGlnAlaGluGlyLysLeuIleThrHisArgSerAlaPheSerArgAla 139
 QY 553 TCACAGATCAACGCTTCCTGGGCTAGTCCCGAGTACCTATGACCTATGCTGTGAGC 612
 Db 140 SerValIleGlnAlaPheLeuGlySerAlaProGlnLeuThrLeuGlnLeuTyrIleSer 159
 QY 613 CTGATCTCTGACAGAGTTCCCTGGGGTAGAGTGTGTGTAATGGTATTTTCCCTGGTATCT 672
 Db 160 ValMetGlnGlnAspValThrValGlyArgSerLeuLeuMetThrIleSerLeuLeuSer 179

QY 673 GTCACTATGGGGCCACCCTTTTGCAATATGTTGGTATCCAGATCAAGTACGATGACTAC 732
 Db 180 IleValTyrGlyAlaLeuArgCysAsnIleLeuAlaIleLysIleLysTyrAspGluTyr 199
 QY 733 AAGATTCCCTTGGCCACCTAGAAAGTCTCTGTCATCACCATCTGGCGGACATTTGGAGATC 792
 Db 200 GluValLysValLysProLeuAlaTyrValCysIlePheLeuTyrArgSerPheGluIle 219
 QY 793 ACTTCCCGCTCTCGATTCTGGTCTCTTCTCAGCCACTTTGAATTTGAAGCTCTGCCC 852
 Db 220 AlaThrArgValValValLeuValLeuPheThrSerValLeuLysThrTrpValValVal 239
 QY 853 TTCTAGTGTCAACTTCTCTGATCATCTCTTGTAGCCCTCGATTAACTTCTGGAGAGT 912
 Db 240 IleLeuIleAsnPhePheSerPhePheLeuTyrProTrpIleLeuPheTrpCysSer 259
 QY 913 GGTGCCAGATGCCCAATAACATTGAGAAAACTTACGCCGGGTCCGACACTCTGTGGTGC 972
 Db 260 GlySerProPheProGluAsnIleGluLysAlaLeuSerArgValGlyThrThrIleVal 279
 QY 973 CTGATTTCACTACCATCTCTATGCTGGCATCAACTTCTCTTGTGCTCAGCTTTGCGAG 1032
 Db 280 LeuCysPheLeuThrLeuLeuTyrThrGlyIleAsnMetPheCysTrpSerAlaValGln 299
 QY 1033 TTGAGGTGGCAGACAGATCTCTGTCACAAAGGGCAGAACTGGGGACATATGGGCTG 1092
 Db 300 LeuLysIleAspSerProAspLeuIleSerLysSerHisAsnTrpTyrGlnLeuLeuVal 319
 QY 1093 CACTATAGTGTGAGTGGTAGAGAATGTGATCATGTCTTGGTTTAAAGTCTTTTGA 1152
 Db 320 TyrTyrMetIleArgPheIleGluAsnAlaIleLeuLeuLeuLeuTyrTyrLeuPheLys 339
 QY 1153 GTCAAGTGTACTGAAATCTACTGTTCATCTCTTGTGCTTGCAGCTCATTTATCTTAT 1212
 Db 340 ThrAspIleTyrMetTyrValCysAlaProLeuLeuValLeuGlnLeuLeuIleGlyTyr 359
 QY 1213 CTGATTTCCATTGCTTCATGCTCTCTTTTCCAGTACTTGCATCCATTCGGCTCCTC 1272
 Db 360 CysThrAlaIleLeuPheMetLeuValPheTyrGlnPheHisProCysLysLysLeu 379
 QY 1273 TTCACCCATATAGTAGTAGAC-----TACCTCCATTGTGTCTGC-----TGT 1314
 Db 380 PheSerSerSerValSerGluGlyPheGlnArgTrpLeuArgCysPheCysTrpAlaCys 399
 QY 1315 CACCAG 1320
 Db 400 ArgGln 401

RESULT 2

XKRY_HUMAN STANDARD; PRT; 159 AA.
 ID XKRY_HUMAN
 AC O14609;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Testis-specific XK related protein Y.
 GN XKRY.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RS SEQUENCE FROM N.A.

RX MEDLINE=98022381; PubMed=9381176;
 RA Lahn B.T., Page D.C.;
 RT "Functional coherence of the human Y chromosome.";
 RL Science 278:675-680(1997).
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -I- TISSUE SPECIFICITY: TESTIS-SPECIFIC.
 CC -I- SIMILARITY: SOME, TO MEMBRANE TRANSPORT PROTEIN XK.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch.)

CC -----
 DR EMBL; AF000997; AAC51844.1; -;
 DR GenBank; HGNC:18571; XKRY.
 DR MIM; 400015; -;
 KW Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 SQ SEQUENCE 159 AA; 18083 MW; B9B81DD842DEFOAB CRC64;

Alignment Scores:
 Pred. No.: 1.23e-08 Length: 159
 Score: 197.50 Matches: 52
 Percent Similarity: 58.55% Conservative: 37
 Best Local Similarity: 34.21% Mismatches: 50
 Query Match: 7.77% Indels: 13
 DB: 1 Gaps: 3

US-09-768-781-2 (1-1389) x XKRY_HUMAN (1-159)

QY 642 AGTTGTGCTAAGTATTTCCCTGCTATCTGTCACCTATGCGGGCCACCTTTGCAATAT 701
 Db 6 SerileAlaAspIlePheProLeuIleSer-CysValGlyAlaIleHisCysAsnII 25
 QY 702 GTTGGCTATCCAGATCAAGTACGATGACTACAGATTCGCTTGGGCCACTAGAACTCCT 761
 Db 25 eLeuAlaIleArgThrGlyAsnAspPheAlaIleLeuValIleLysLeuII 45
 QY 762 CTGCATCACCATTCTGGCGGACATTGAGATCACTTCGCGCTCCTGATTCGTGCTCTT 821
 Db 45 eTyxLeuMetIleTrpHisSerLeuValIleLeuSerProValThrLeuAlaPhePh 65
 QY 822 CTCAGCCATTGAAATTCAGGCTGTGCGCTTCCCTAGTCTCAACTTCTGTGATCATCT 881
 Db 65 eProAlaSerLeuLysGlnGlySerLeuHisPheLeuLeuIleIleTyxPheValLeuLe 85
 QY 882 CTTTGAGCCCTGGATTAAGTTCTGGAGAGTGTGGCCAGATGCCCAATAACATTGAGAA 941
 Db 85 uLeuThrProTrpLeuGluPheSerLysSerGlyThrHisLeuProSerAsn-ThrLysI 105
 QY 942 AAATTTCAGCGGTGCGGCACTCTGTGG-----TCCTGATTCAGTCCACAT 989
 Db 105 leIleProAla-----TyrTrpValSerMetAspAla-TyrLeuAsnHisAla 120
 QY 990 CTTCTATGCTGCATCAACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1049
 Db 121 SerIleCysCysHisGln-PheSerCysLeuSerAlaValLysLeuGlnLeuSerAsnG 140
 QY 1050 AGATCTCGCGCAAGGGCAGAACTGG 1077
 Db 140 uGluLeuIle---ArgAspThrArgTrp 148

RESULT 3

ID DTPT LACHE STANDARD; PRT; 497 AA.
 AC 007380.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DI-/tripeptide transporter.
 GN DTPT.
 OS Lactobacillus helveticus.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1587;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCDO 2712;

RX MEDLINE=97316430; PubMed=9172341;
 RA Nakajima H., Hagting A., Kunji E.R.S., Poolman B., Konings W.N.;
 RT "Cloning and functional expression in *Escherichia coli* of the gene
 encoding the di- and tripeptide transport protein of *Lactobacillus*
 RT *helveticus*.";
 RL Appl. Environ. Microbiol. 63:2213-2217 (1997).
 CC -|- FUNCTION: PROTON-DEPENDENT UPTAKE OF DI- OR TRI-PEPTIDES.
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -|- SIMILARITY: BELONGS TO THE PTR2 FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch.)

CC -----
 DR EMBL; U77486; AAC45382.1; -;
 DR InterPro; IPR000109; PTR2.
 DR Pfam; PF00854; PTR2; 1.
 DR TIGRFAMs; TIGR00923; 2A1701; 1.
 DR PROSITE; PS01022; PTR2_1; 1.
 DR PROSITE; PS01023; PTR2_2; 1.
 KW Peptide transport; Transport; Transmembrane.
 FT TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 84 104 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 FT TRANSMEM 227 247 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.
 FT TRANSMEM 372 392 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 SQ SEQUENCE 497 AA; 55469 MW; 69C064FF9A0A26A5 CRC64;

Alignment Scores:
 Pred. No.: 0.324 Length: 497
 Score: 108.50 Matches: 85
 Percent Similarity: 36.76% Conservative: 65
 Best Local Similarity: 20.83% Mismatches: 133
 Query Match: 4.27% Indels: 125
 DB: 1 Gaps: 18

US-09-768-781-2 (1-1389) x DTPT_LACHE (1-497)

QY 310 GATCTAGCCAAAGATAACGGCTATCATTTATTCATCTCTTGGGACCTGTT 369
 Db 110 AspProArgAspAlaGlyPheSerIlePheValPheGlyIleAsnLeuGlySerIle 129
 QY 370 ATCAGA----- 375
 Db 130 IleAlaProTrpLeuValProTrpAlaAlaGlnGlyPheGlyValHisIlePheGlySer 149
 QY 376 -----TGTGGAGGCCATGATT 393
 Db 150 GlnLeuAsnPheHisAlaGlyPheSerLeuAlaAlaValGlyMetPhePheGlyLeuVal 169
 QY 394 AGTACCTCACACTGTGGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 453
 Db 170 GlnTyxValLeuGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 186
 QY 454 AGAAGATCTAATAGATGCGGAGGAGGTG---CTGATAGATGGGAGGTGGGCACTCC 510
 Db 187 AsnAspProIleAspLysGlyAspLeuLeuAsnValIleLysTrpValIleIleIle 206
 QY 511 ATCCGAGCCCTGGCTATGCACCCCAATGCCACAAAGTATGTCACAGACCTTC 570
 Db 207 IleAlaIleValAla-----IleLeuAlaAla 215

```
QV 571 CTGGGCTCAGTGCCTCCAGCTGACCTATCAGCTCTATGTGACGCTG-----ATCTCT 621
Db 216 MetAlaGlyValGlyGlnLeuSerValAspAenValIleThrLeuLeuLeuAala 235
QV 622 GCAGAGTTCCTCCCTGGGTAGAGTTGCTAATGTATTTTCC-----CTGGTATCTGTC 675
Db 236 IleAlaLeuProIleTyrTyrPheValMetMetPheArgSerSerLysValThrLysIle 255
QV 676 ACCTATGGGGCCACCTTTTGCATATG-----702
Db 256 GluLeuGlyIleHisLeuLeuProValSerLeuLysAenArgLeuPhePheLysGly 275
QV 703 -----TTGCTATCCAGATCAAGTACGACTACGACTACAAG 735
Db 276 TyrLysArgLeuLysGlnIleLeuGlnLeuAlaIleLysArgGlnSerPheIle 295
QV 736 ATTGGCTTGGGCGACATAGAGTCTCTGATCACCATCTGCGGCGACATTTGGAGTCACT 795
Db 296 IleLeuIleAlaLeuIleIleMetAlaSerIleLeuIleProAenLysValIleIleAala 315
QV 796 TCCCGCTCTCTGATCTGCTGCTCTTC---TCAGCCACTTGAATGAAGGCTGTGCC 852
Db 316 LysHisLeuLeuLysValLeuValPheTyrTrpIleGlyLeuAenLeuIlePro 335
QV 853 TTC-----CTAGTGTCAACTCTCTCATCATCTCTTTGAGCCCTCGATTAAAG----- 900
Db 336 PheSerThrPheValLeuSerPhe-----LeuPheLeuAspTyrIleLysHisMet 352
QV 901 TTCTGGAGAGTGTGCTCCAGATGCCCAATAACATTGAGAAAC----- 945
Db 353 PheLysLysGluGlyGlnAlaLysThrLysGlyLysSerArgIleHisGly 372
QV 946 -----TTCAGCCCG-----GTCGGCATCTGCTGGTGTCTG 975
Db 373 IleGluIleProLeuPheLeuArgGlnLeuIleAenIlePheThrLeuIleIleLeu 392
QV 976 ATTTCAGTACCATCTCT---TATGCTGGCATCACTTCTCTTGC----- 1017
Db 393 GluGlyGluThrLeuPheAspGluAsnGlyValGluValAenIleAlaGluHisProVal 412
QV 1018 -----TGCTCAGCTTTGCGAGTTGAGTTGGCAGACAGATCTCTGCGACAAAGGCGAG 1071
Db 413 GlnGlyTyrThrGluLeuAenIleAenLeuAenLysAspSerIleAspLeuTrpAla 432
QV 1072 AACTGGGACATATGGGCTGCGACTATAGTGTGAGGTTGGTAGAATGTGATCATGGTC 1131
Db 433 AspTrp-----IleGlnSerVal----- 438
QV 1132 TTGGTTTAAAGTTCTTTGGAGTGAAGTCTTACTGAATCTACTGCTCCTTCTTCTGTC 1191
Db 439 -----AlaLysTyrLeuLeuAenIleMetTyrThrAlaAspVal 451
QV 1192 TTGAGCTCATTATTGTTATCTGATTTCCATTCGCTTCATGCTCTTCTTCCAGTAC 1251
Db 452 IleValIleIleIlePheTyrLeuValLysMetAlaLeuLeuTrpTrpAlaTrpSerTyr 471
QV 1252 TTGCATCCATTTGGCTCACTCTTC 1275
Db 472 Ile---ProLeuSerThrValPhe 478

RESULT 4
YBIO_ECOLI
ID YBIO_ECOLI STANDARD; PRT; 741 AA.
AC P75783; Q9R7S0; Q9R7S2; Q9ZBC5;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ybio.
GN YBIO OR B0808.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
```

```
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:11453-1474 (1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa K., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
Sampai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0003 FAMILY.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
frameshift in position 77.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000183; AAC73895.1; ALT_INIT.
DR EMBL; D90717; BAA35474.1; -.
DR EMBL; D90718; BAA35479.1; ALT_FRAME.
DR EMBL; D90718; BAA35480.1; ALT_FRAME.
DR EcoGene; EGI3320; Ybio.
DR InterPro; IPR001880; MSion channel.
DR Pfam; PF00924; MS_channel_1.
DR PROSITE; PS01246; UPF0003; FALSE_NEG.
DR KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 268 288 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 466 486 POTENTIAL.
FT TRANSMEM 509 529 POTENTIAL.
FT TRANSMEM 533 553 POTENTIAL.
FT TRANSMEM 562 582 POTENTIAL.
FT TRANSMEM 608 628 POTENTIAL.
FT TRANSMEM 662 682 POTENTIAL.
FT DOMAIN 25 30 POLY-THR.
FT DOMAIN 195 202 POLY-LEU.
SQ SEQUENCE 741 AA; 81903 MW; 69B0A27678976DB5 CRC64;

Alignment Scores:
Pred. No.: 0.347 Length: 741
Score: 108.50 Matches: 97
Percent Similarity: 37.88% Conservative: 92
Best Local Similarity: 19.44% Mismatches: 177
Query Match: 4.27% Indels: 133
DB: 1 Gaps: 25

US-09-768-781-2 (1-1389) x YBIO_ECOLI (1-741)
```


Pred. No.:	0.597	Length:	521
Score:	104.50	Matches:	72
Percent Similarity:	37.77%	Conservative:	70
Best Local Similarity:	19.15%	Mismatches:	139
Query Match:	4.11%	Indels:	95
DB:	1	Gaps:	16

US-09-768-781-2 (1-1389) x YT25_CABEL (1-521)	
Qy	385 GCCATGATTAAAGTACCTCACACTGTGGAAAGAAAGAGGAGGAGGAGCCCTATTGTCAGC 444
Db	168 AlaAlaLeuAspPheLeuMetMetTTP-----TyrThrTrp 179
Qy	445 CTCACCCGAAGAAGATGCTAATAGATGGCGAGGAGGTGCTGATAGATGGAGGTGGGC 504
Db	180 AsnAspArgLysIleAenTyAspLeuAsnSerGlnTyrLeuLysValGluGlnHis 199
Qy	505 CACTCCATCCGGACCCCTGGCTATGTCACCGCAATGCCTACAAACGATGTCACAGATCCAA 564
Db	200 HisSerLeuMetAlaValSerLeuAsn-----SerLeuLeuGln 212
Qy	565 GCCTTCCTGGGCTCAGTGCACCTGACCTCATGAGCTCATGTGAGCCTGATCTCTGCA 624
Db	213 LeuPheValThrSerIleTyrAlaIleSerMetPheValLeuAlaAsnMetSerMetThr 232
Qy	625 GAGGTTCCCTCGGTAGAGTT---GTGCTAATGGTATTTTCCTGGTATCTGTCACTTAT 681
Db	233 AsnProGluLeuGlyAsnAlaAenLeuLeuArgTTPPheTyrAlaPheLysValLeuTyr 252
Qy	682 GGGGCCACCCCTTTGCAAT-----ATGCTGGCTATCCAGATCAAGTACCAT 726
Db	253 ArgArgSerValPheHisIleSerThrHisGlnLeuLeuPheMetSerIleValTyrAla 272
Qy	727 GACTPAC---AAGATTCCGCTTGGCCACCTAGAGTCTCTGTCATCACCATCTGGCGGACA 783
Db	273 AspPheHisAlaIleSerTyrGlyPheLeuGln-----AlaIleSerLeuTyrArgSer 290
Qy	784 TTGGAGATCATCTCCGCTCTGATTCGGTGCTCTTCTCAGCCACTTTGAAATTGAAG 843
Db	291 ValValPheTyrAspAsp---ProValMetIleMetPheThrGly----- 304
Qy	844 GCTGTGCCCTTCCTAGTGTCAACTCTCTGATCATCTCTTGGAGCCCTGGATTAAAGTTC 903
Db	305 -----TyrGluCysTyrAlaIleThr 311
Qy	904 TGGAGAAAGTGTTGCC-----CAGATGCCCAATAACATTGAG 939
Db	312 TrpThrValGlyAlaAlaLysMetLeuMetSerPheIleGlnThrAlaLeuThrIleGlu 331
Qy	940 AAAAACTTCAGCCGGGTCCGCACTCTGGTG----- 969
Db	332 ArgIleIleAspArgIleIleProLeuValProLysLeuLysProPheLysArgGlnGly 351
Qy	970 GTCTCGATTTCAGTCACCATCTCTATGTGGCATCAACTCTCTGTGTGGTCACTTTG 1029
Db	352 LeuPheLeuAsnAlaPheAlaLeuThrAlaGlyIleSerThrThrIleTyrSerTyrSer 371
Qy	1030 CAG-----TTGAGGTTGGCA-----GAC 1047
Db	372 GluGlyProThrGlyTyrLysLeuAlaSerCysPheMetGlnLysAspIleProLeuAsp 391
Qy	1048 AGAGATCTCGTCGACAAAGGGCAGAACTGGGACATATGGGCTGCACCTATATAGTGTGAGG 1107
Db	392 ArgValLeuTyrThrLeuGlyMet-----TyrPheAlaLeuSerLeuLeuCysLeu 408
Qy	1108 TTGTAGAGAAATGTGATCATGTCCTTGGTTTTT-----AAGTTCCTTGGAGTG 1155
Db	409 LeuAlaAsnLeuThrIleIlePheSerIlePheLysSerSerLysLysSerPheAsnLeu 428
Qy	1156 AAAGTGTTACTGAAT-----TACTGTTCATCTTCATTGCTCC 1191
Db	429 LysIleArgPheAsnPheGlnGluValLysAsnSerSerPheAlaValSerIleLeuSer 448

Pred. No.: 1.46 Length: 459
 Score: 100.50 Matches: 79
 Percent Similarity: 36.84% Conservative: 75
 Best Local Similarity: 18.90% Mismatches: 131
 Query Match: 3.95% Indels: 133
 DB: 1 Gaps: 20

US-09-768-781-2 (1-1389) x NU4M_CERSI (1-459)

```

QY 238 ATGACATACACC-----TTTCTTCTTTATGTTTTCATCCATTATGGTC 282
DB 108 MetThrPheThrAlaThrGluLeuLeuPhePheTyrIleLeuPheGluAlaThrLeuIle 127
QY 283 CAGTTGACCTCAATTTTGTCCACAGAGATCTAGCCAAAGATAAACCCTGATCA- 336
DB 128 Pro---ThrLeuIleIleThrArgTrpGlyAsnGlnThrGluArgLeuAsnAlaGly 146
QY 337 ---TTATTATGATCAATCTCTCTGGGA-----CCTCTATCAGATGTTTGGAGGCC 387
DB 147 PheTyrPheLeuPheTyrThrLeuThrGlySerLeuProLeuLeu----- 161
QY 388 ATGATTAACTACCTACACTGTGTGGAAGAAAGAGAGAGAGAGAGCCCTATGTCAGCCTC 447
DB 162 ----- 164
QY 448 ACCGAAAGAAAGATGCTAATAGATGGCAGAGGTGCTGATAGATGGAGGTGGGCCAC 507
DB 165 ThrHisIleGlnAsnLeuThrGlySerLeuAsnLeuLeu-----IleGlnTyr 181
QY 508 TCATCCGACCCCTG---GCTATGACCGCAAGTACCAAAAGTATGTACAGATCCAA 564
DB 182 SerThrGlnThrLeuProSerSerTrpSerAsnAlaPheLeuTrpLeuAlaCysMetMet 201
QY 565 GCCTCTCTGGCTCAGTGCCTCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCGCA 624
DB 202 AlaPheMetValMetProLeuTyrGlyLeuHisLeuTrpLeuProLysAlaHisVal 221
QY 625 GAGGTTTCCCTCG---GGTAGAGTGTGCTAATGGTATTTTCCCTGGTATCTGTACCTAT 681
DB 222 GluAlaProIleAlaGlySerMetValLeuAlaIleLeuLeuLysLeuGlyGlyTyr 241
QY 682 GGGGCCACCCCTTTGCAATATGTTGGCTATCCAGATCAAGTACGATCAAGATTCGC 741
DB 242 GlyMetLeuArgIleThrMetIle----- 249
QY 742 CTGGGCCACTA-----GAAGTCTCTGCATCACCATCTCGGGGACA 783
DB 250 LeuAsnProLeuThrSerTyrMetAlaTyrProPheLeuMetLeuSerLeuTrpGlyMet 269
QY 784 TTGGAGATCACTTCCCGCTCTGATTTCTGTGCTCTCTCAGCCACTTTGAAATTGAAG 843
DB 270 IleMetThrSerSerIleCysLeuArgGlnThrAspLeuLysSerLeuIleAlaTyrSer 289
QY 844 GCTGTGCCCTCTCTAGTGTCTCACTCTCTGATCATCTCTTTGAG---CCCTGGATTAAAG 900
DB 290 SerValSerHisMetAlaLeuValIleValAlaValIleGlnThrProTrp---Ser 308
QY 901 TTCTGGAGAGTGGTCCCATG- 924
DB 309 TyrMetGlyAlaThrAlaLeuMetIleAlaHisGlyLeuThrSerSerValLeuPheCys 328
QY 925 -----CCCAATAACATTGAGAAAACTTCAGCCGG-----GTC 957
DB 329 LeuAlaAsnSerAsnTyrGluArgThrHisSerArgThrMetIleLeuAlaArgGlyLeu 348
QY 958 GGCATCTGTGGTCTCTGATTTTCAGTCACCATCTCTATGCTGGCATCACTTCTCTTGC 1017
DB 349 GlnThrLeuLeuProLeuMetAlaMetTrpTrpLeuLeuAlaSerLeuThr----- 365
QY 1018 TGGTCAGCTTTGCGAGTTGAGTTGGCAGACAGATCTCGTCGACAAAGGGCAGAACTGG 1077
DB 365 ----- 365

```

```

QY 1078 GGACATATGGCGCTGCACATATAGTGTGAGTGTGGTAGAAGATGTGATCATGGTCTTG--- 1134
DB 366 ---AsnLeuAlaLeuProThrIleAsnLeuValGlyGluLeuPheValValMetSer 384
QY 1135 GTTTTAAAGTCTTTGGAGTGAAGCTTACTGTAATTACTGTCTCATCTCTTGATTCCTTG 1194
DB 385 SerPheSerTrpSerAsnIleThrIleIle-----LeuMetGlyThr 398
QY 1195 CAGCTCATATTATGCTTATCTGATTTCATTTGGCTTTCATGCTCCCTTTCTCCAGTACTTG 1254
DB 399 AsnIleIleIleThrAlaLeuTyrSerLeu---TyrMetLeuIleThrThrGln----- 415
QY 1255 CATCCATTGCGCTCACTCTTCACCCATAATGTAGTAGACATACCTCCATTGTGTCTGTGT 1314
DB 416 -----ArgGlySerTyrThrHisHis----- 422
QY 1315 CACCAGCACCTCGGACCGAGGTGAGAACTCAGAGCCACCCCTTTGAGACTGAA 1368
DB 423 -----IleAsnAsnIleLysProSerPheThrArgGlu 433

```

RESULT 9

```

YNSI_YEAST STANDARD; PRT; 551 AA.
ID YNSI_YEAST
AC P53730;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 62.7 kDa protein in SEC12-SSK2 intergenic region.
GN YNR030W OR N3265.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]_TaxID=4932;
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO C.ELEGANS ZC513.5 AND S.POMBE SPBC1734.12C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z71645; CAA96310.1; -.
DR SGD; S0005313; YNR030W.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
FT TRANSMEM 203 223 POTENTIAL.
FT TRANSMEM 228 248 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 366 386 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
SQ SEQUENCE 551 AA; 62672 MW; B08A94BBF260502F CRC64;

```

Alignment Scores:

```

Pred. No.: 1.82 Length: 551
Score: 99.50 Matches: 72
Percent Similarity: 32.68% Conservative: 57
Best Local Similarity: 18.23% Mismatches: 115
Query Match: 3.91% Indels: 151
DB: 1 Gaps: 15

```

US-09-768-781-2 (1-1389) x YN8I YEAST (1-551)

259	QY	TTTATGTTTTTCATCCATTATGTGTCAGTCTGACCCCTCATTTTGTGCCAGAGACTAGCC	318
75	Db	TyrLeuTyrLeuSerLeuLeuGlnThrSer	85
319	QY	AAAGATAAACCGCTATCATTTATTCATCATTAATCCTCTTGGGACCTGTT	369
86	Db	ArgProThrSerIleAspValGlnLeuValValArgGlyIleValGlyLeuThr	103
370	QY	ATCAGATCTTTGGAGGCCATGATTAAAGTACCTCCTCAC	405
104	Db	AsnGlyLeuSerPheIleTyrLeuLysAsnCysLeuGlnAspMetPheAspGluIleThr	123
406	QY	CTGTGGAAGAAAGAGGAGGAGGAGCC	441
124	Db	GluLysLysGluAsnGluAspLysAspIleTyrIleTyrAspSerAlaGlyThr	143
442	QY	AGCCTC	447
144	Db	TrpPheLeuLeuPheLeuIleGlySerPheHisLeuMetPheTyrSerThrArgThrLeu	163
448	QY	ACCGAAAGAGATGCTAATAGATGCGAGGAGGTGCTAGTAATGGAGGTGGCCAC	507
164	Db	ProAsnPheValMetThrLeuProLeuThrAsnValAlaLeuGlyTrp	179
508	QY	TCCATCCGACCTCGCTATGCACCGCAATGCTACAAACGTATGTCACAGATCCAAGCC	567
180	Db	ValLeuLeuGlyArgTyrAsnAla	189
568	QY	TTCTGGCTCAGTCCCGCAGCTGACCTATGAGCTCTATGTGAGCTGATCTCGACAG	627
190	Db	PheLeuSerAlaLeuValAlaIleValPheArgLeuGluValSerAlaLeuSerAlaGly	209
628	QY	GTTCCTCCCTGGTAGATTGTGCTAATGGTATTTTCCTGGTATCTGTC	675
210	Db	IleAlaLeuPheSerValIlePheLysIleSerLeuPheAspAlaIleLysPheGly	229
676	QY	ACCTATGGGCGCCACTTTGGCAATATGTTGGCTATTCAGATCAAGTACGATGACTAC	732
230	Db	IlePheGlyLeuGlyLeuGlySer	247
732	QY	TrpGlnGluTrpCysLeuProGluValAspGlyPheLeuPheAsnValValAlaGlyTyr	267
733	QY	NAGATTCGCTTGGCCACTAGAAAGTCCCTCGATCACCATCTGGCGGACATTG	786
268	Db	AlaSerIleTrpGlyValGluProValThrAlaTyrPheThrHisTyrLeuArgMetMet	287
787	QY	GAGATCACTTCCCGCTCTGATTCGTGGTCTTCTCA	837
288	Db	PheMetProThrValLeuLeuLeuAsnTyrPheGlyTyrLysLeuAlaProAlaLys	307
838	QY	TTGAGGCTGTGCCCTTCCTAGTGTCAACTCTTCCTGATCATCTCTC	894
308	Db	LeuLysIleValSerLeuAlaSerLeuPheHisIleIleValLeuSerPheGlnProHis	327
895	QY	ATTAAAGTCTGGAGAAGTGGTGGCCAGATGCCCAATAACATTGAGAAAACCTCAGCGGG	954
328	Db	LysGluTrpArg	331
955	QY	GTCCGCACTCTGGTGGTCTGATTTTCAGTCACCATCTCTATGCTGGCATCACTTCTCT	1014
332	Db	PheIleIleTyrAlaValProSerIleMetLeuLeuGlyAlaThrGlyAla	348
1015	QY	TGCTGTGTGAGCTTTGCAAGTTTGGCAGACAGAGATCTCGTGCACAAAGGCGAGAAC	1074
348	Db		348
1075	QY	TGGGGACATATGGCCCTGCATATAGTGTGAGTGGTAGAATGTGATCATCGTCTTG	1134
349	Db	AlaHisLeuTrpGluLeuMetLysValLysIleThrAsnValLeu	364

Qy	1135	GTTTTAAAGTCTTTGGAGCTGAAGGTGTACTCGAATTACTGTCATCTCTCGATTGCCTTG	1194
Db	365	-----CysLeuAlaIleLeuProLeu	371
Qy	1195	CAGCTCATTATGTCTTATCTGATTTCCATTGCTTCATGCTCCTT	1239
Db	372	SerIleMetThrSerPhePheIleSerMetAlaPheLeuTyrIle	386

RESULT 10

ATPI_OCHNE STANDARD; PRT; 233 AA.
 ID ATPI_OCHNE Q40607;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP synthase A chain precursor (EC 3.6.3.14) (Subunit IV).
 GN ATPI.
 OS Chlorosphaera neapolitana.
 OG Chloroplast.
 OG Eukaryota; Haptophyceae; Isochrysidales; Ochrosphaera.
 OX NCBI_TaxID=35137;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CCMP 593;
 RA Huss V.A.R., Tietze A.C., Julius C.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -I- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out)
 CC -I- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
 CC thylakoid membrane.
 CC -I- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collabora-
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no
 CC modified and this statement is not removed. Usage by and for commer-
 CC entities requires a license agreement (See <http://www.ebi-sib.ch/annou>-
 CC or send an email to license@ebi-sib.ch).

Qy 712 CAGATCAAGTACGATCACTACAGATTCCGCTTGGCCACTAGAA-----GTC 759
 Db 15 -----TyrTrpGluileGlyLeuGluValHisGlyGlnVal 27
 Qy 760 CTCGTCATCACCATCTGGGGACATGGAGATCACTCCCGCTCTGATCTGGTGCTC 819
 Db 28 LeuLeuileThr-----TrpLeuValLeuAlaileLeuThrLeuAlaileLeu----- 44
 Qy 820 TTCTCAGCCACTTTGAAATTTGAAGGCTGTGCGCTTCTAGTGCTCACTTCCTGATCATC 879
 Db 45 -----GlyThrLeuLeuGluGlnValProlyGlyValGlnAsnPheLeuGluSer 62
 Qy 880 CTTCTTGAG-----CC 891
 Db 63 ValPheGluTyrrValSerGlyileAlaLysAspGlnileGlyGluTyrrHisTyrrArgPro 82
 Qy 892 TGGATTAAAGTTC-----TGGAGAATGGT 915
 Db 83 TrpValProPheValGlyThrLeuPheLeuPheIlePheValAlaAsnTrp-----LeuGly 101
 Qy 916 GCCCAGATGCC-----AAT 930
 Db 102 AlaLeuileProTrpLysLeuileHisLeuProGluGlyLeuAlaalaProThrAsn 121
 Qy 931 AACATTGAGAAAAACTTCAGCGGGTGGCAGCTCTGGTGTCTGATTTTCAGTCACCATC 990
 Db 122 AspIleAsnThrValala-----LeuSerLeuLeuThrSerIleSerTyrr 137
 Qy 991 CTCATTGTCGATCAACTTCTTCTGCTGTCAGCTTTGCGAGTTGAGTTGGCAGACAGA 1050
 Db 138 PheTyrrAlaGlyPheLys----- 143
 Qy 1051 GATCTCGTCGACAAAGGCGAGAACTGGGACATATGGCTGCATATAGTGTGAGTTG 1110
 Db 144 -----GluLysGlyLeuGlyPhePheAlaArgTyrr 153
 Qy 1111 GTAGAG-----AATGTCATC-----ANG 1128
 Db 154 IleSerProThrProIlePheLeuProIleAsnileLeuGluAspPheThrLysProLeu 173
 Qy 1129 GCTTTGGTTTTTAAGTCTTTTGA-----GTGAAAGTGTTA 1164
 Db 174 SerLeuSerPheArgLeuPheGlyAsnileLeuAlaAspGluileValValSerValLeu 193
 Qy 1165 CTGAATTACTGTCATCTTCTGATGCTGAGCTCATTTATGCTTATCTGATTTCCATT 1224
 Db 194 CysLeuLeuValProLeuLeuileProLeuProValMetValLeuGlyilePheAlaSer 213
 Qy 1225 GCGTTCATGCTCTTTCTTCCAGTACTTG 1254
 Db 214 SerValGlnAlaLeuValPheSerThrLeu 223
 RESULT 11
 ID ACHO HUMAN STANDARD; PRT; 458 AA.
 AC Q05901, Q15827;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neuronal acetylcholine receptor protein, beta-3 chain precursor.
 GN CHRN3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Substantia nigra;
 RX MEDLINE=97062879; PubMed=8906617;
 RA Elliott K.J., Ellis S.B., Berckhan K.J., Urrutia A.,
 RA Chavez-Noriega L.E., Johnson E.C., Velicelebi G., Harpold M.M.;
 RT "Comparative structure of human neuronal alpha 2-alpha 7 and beta

2-beta 4 nicotinic acetylcholine receptor subunits and functional expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and beta 4 subunits.";
 J. Mol. Neurosci. 7:217-228 (1996).
 [2]
 RN SEQUENCE FROM N.A.
 RP Tissue=Pons;
 RX MEDLINE=97162233; PubMed=9009220;
 RA Groot Kormelink P.J., Luyten W.H.M.L.;
 RT "Cloning and sequence of full-length cDNAs encoding the human neuronal nicotinic acetylcholine receptor (nAChR) subunits beta3 and beta4 and expression of seven nAChR subunits in the human neuroblastoma cell line SH-SY5Y and/or IMR-32.";
 FEBS Lett. 400:309-314 (1997).
 [3]
 RN SEQUENCE FROM N.A.
 RP Keddache M., Durrner M., Greenberg D.A.;
 RT "Genomic structure and mutation analysis of the CHRN3 gene.";
 Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE OF 36-458 FROM N.A.
 RP Tissue=Brain stem;
 RX MEDLINE=93390761; PubMed=7690916;
 RA Willoughby J.J., Ninkina N.N., Besch M.M., Latchman D.S., Wood J.N.;
 RT "Molecular cloning of a human neuronal nicotinic acetylcholine receptor beta 3-like subunit.";
 Neurosci. Lett. 155:136-139 (1993).
 RL Neurosci. Lett. 155:136-139 (1993).
 CC -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA MEMBRANE.
 CC -!- SUBUNIT: NEURONAL ACHR SEEMS TO BE COMPOSED OF TWO DIFFERENT TYPE OF SUBUNITS: ALPHA AND BETA.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U62438; A840116.1; --
 DR EMBL; Y08417; CA69694.1; --
 DR EMBL; AF140765; AAD33063.1; --
 DR EMBL; AF140760; AAD33063.1; JOINED.
 DR EMBL; AF140761; AAD33063.1; JOINED.
 DR EMBL; AF140762; AAD33063.1; JOINED.
 DR EMBL; AF140763; AAD33063.1; JOINED.
 DR EMBL; AF140764; AAD33063.1; JOINED.
 DR EMBL; X67513; CAA47851.1; --
 DR PIR; S25587; S25587.
 DR Genew; HGNC:1963; CHRN3.
 DR MIM; 118508; --
 DR InterPro; IPR00188; GABAA_receptor.
 DR InterPro; IPR001175; Neur_channel.
 DR Pfam; PF02931; Neur_chan_LBD; 1.
 DR Pfam; PF02932; Neur_chan_memb; 1.
 DR PRINTS; PR00252; NRIONCHANNEL.
 DR TIGRfam; TIGR00860; LIC; 1.
 DR PROSITE; PS00236; NEUOTR_ION_CHANNEL; 1.
 KW Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
 KW Transmembrane; Multigene family.
 FT SIGNAL 1 24
 FT CHAIN 25 458
 FT NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN.
 FT BETA-3 CHAIN.
 FT EXTRACELLULAR (POTENTIAL).
 FT POTENTIAL.
 FT TRANSMEM 233 257
 FT TRANSMEM 265 282
 FT TRANSMEM 299 320
 FT TRANSMEM 321 428
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN

```

FT TRANSMEM 429 447 POTENTIAL.
FT DISULFID 153 167 BY SIMILARITY.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 36 38 LQF -> EWK (IN REF. 3).
SQ SEQUENCE 458 AA; 52728 MW; D6E919E53CBD21F8 CRC64;

Alignment Scores:
Pred. No.: 1-94 Length: 458
Score: 99.00 Matches: 45
Percent Similarity: 45.71% Conservative: 35
Best Local Similarity: 25.71% Mismatches: 63
Query Match: 3.89% Indels: 8
DB: 1 Gaps: 8

US-09-768-781-2 (1-1389) x ACHO_HUMAN (1-458)
QY 256 TTCCTTTATGTTTCATCCATTATGGTCAGTGCACCTCATTTTGTGTC--- 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 TyrLeuPheLeuMetilePheValThrLeuSerileleValThrValPheValle 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 -----CACAGATCTAGCAAGATAAACCGTATCATATTATTATGATCTAATC 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 AsnValHisArgSerSerThrTyrHisProMetAlaProTrpValLysArgLeu 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 CTCCTGGGA-----CTGTTATCAGATGTTTGGAGGCCATGATTAGTACCTCACACTG 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 PheLeuGlnLysLeuProLysLeuLysCysMetLysAspHisValAspArgTyrSerSer 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 409 TGAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 ProGluLysGluGlu---SerGlnProValValLysGlyLysValLeuGluLysLysLys 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 454 AAGAGATGCTAATAGATGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 378 GlnLysGlnLeuSerAspGlyGluLysValLeuAlaPheLeuGluLysAlaAlaAsp 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 508 TCCATCCGACCTGGCTATGACCGCAATGCTTACAAAGTATGTCACAGATCCAGGCC 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 SerileArgTyrileSerArgHisValLysLysGluHisPheileSerGlnValValGln 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 568 TTCCTGGGCTCAGTGCCTCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTGCAGAG 627
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 AspTrpLysPheValAlaGlnVal----- 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 628 GTTCCCTGGTAGAGTGTGCTAATGTTATTTTCCTGCTATCTGTACCTATGGGGCC 687
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 -----LeuAspArgilePheLeuTrpLeuPheLeuileValSerValThr---GlySer 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 688 ACCCTTGCATATGTTGCTATCCAGATCAAGTACGATGACTAC 732
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 ValLeuPhePheThrProAlaLeuLysMetTrpLeuHisSerTyr 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
VPH1_SCHPO STANDARD; PRT; 805 AA.
AC 013742;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable vacuolar ATP synthase 91 kDa subunit (Vacuolar ATPase 91 kDa subunit).
GN SPAC1688.07C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;

```

```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Spouras J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Horsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: REQUIRED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR
CC ATPASE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF
CC THE ENZYME FOR A SPECIFIC ORGANELLE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. VACUOLAR
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE V-ATPASE 116 kDa SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z98529; CAB11035.1; -
DR InterPro; IPR002490; V_ATPase sub116.
DR Pfam; PF01496; V_ATPase sub a7.1.
KW Hydrogen ion transport; Transmembrane; Glycoprotein.
FT DOMAIN 1 404 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 405 429 1 (POTENTIAL).
FT DOMAIN 430 451 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 452 472 2 (POTENTIAL).
FT DOMAIN 473 526 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 527 547 3 (POTENTIAL).
FT DOMAIN 548 558 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 559 579 4 (POTENTIAL).
FT DOMAIN 580 590 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 591 611 5 (POTENTIAL).
FT DOMAIN 612 624 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 625 645 6 (POTENTIAL).
FT DOMAIN 646 685 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 686 706 7 (POTENTIAL).
FT DOMAIN 707 745 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 746 766 8 (POTENTIAL).
FT DOMAIN 767 805 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 805 AA; 91226 MW; DE92CCBACC66C6CF CRC64;

Alignment Scores:
Pred. No.: 2-58 Length: 805
Score: 98.00 Matches: 97
Percent Similarity: 37.47% Conservative: 78
Best Local Similarity: 20.77% Mismatches: 174
Query Match: 3.85% Indels: 118

```

DB: 1 Gaps: 25

US-09-768-781-2 (1-1389) x VPH1_SCHPO (1-805)

QY 22 TCAGAAAGAACCTCGACAATGAGCAGAGTTTATGAATTCCTGAG-----GAGCCAAAT 75
 DB 362 SerGluGlnProThrTyrrPheArgValAsnLysPheThrGluGlyPheGlnSerIle 381
 QY 76 GTGGATCGGGTTTCA---TCTCTGAGAGAAATGTCATCCGTGGAGGCAACCCCGGATTT 132
 DB 382 IleAspSerTyrrGlyIleAlaThrTyrrArgGluValAsnHisGlyIleValAlaIleVal 401
 QY 133 ACTTTTCA-----TTAGCATCCTTTTCCACCTTTTGTGACTGTGGGAGGCTGCA 186
 DB 402 ThrPheProPheLeuPheAlaIleMetPheGlyAspLeuGlyHisGlyAlaIleMetAla 421
 QY 187 TCTGCTTTGTACATGGTTAGATCATTCGAAAGAAAT-----AGT 225
 DB 422 SerValAlaLeuMetPheValLeuTyrrGluLysThrLeuGlyAlaLysLysAspLeuAsp 441
 QY 226 GAAACTTACTGGATGACATCACCTTTTCTTCTTTATGTTTTCATCCATTTAGTGCAG 285
 DB 442 GluIleValGlyMetValPheTyrrGlyArgTyrrIleValLeuLeuMetGlyLeuPheSer 461
 QY 286 TTGACCTCATTTTTCACAGAGATCTACCCAAAGATAAACCGCTATCATTTATTATG 345
 DB 462 MetTyrrValGlyPheValTyrrAsnAspLeuPhe---SerLysProMetSerIlePheSer 480
 QY 346 CATCTAATCCTCTGGGACCTGTTATCATGATGTTGGAGGCCATGATTAGTACCTCACA 405
 DB 481 Ser-----ArgTrpValTrp 485
 QY 406 CTGTGGAAGAAAGAGGAGGAGGAGGCTTATGTGACGCTTCCACCGAAAGAGAGCTA 465
 DB 486 ProValLysSerGluGluAlaIleAlaArgAlaValGlnValGlyThrTyrrProIleGly 505
 QY 466 ATAGATCGCGAGGAGGTCTGATAGATGGAGGTGGGCCACTCCATCCGAGACCTGGCT 525
 DB 506 IleAspPro-----ThrTrp-----HisSerAlaAspAsnAsnLeu 517
 QY 526 ATGACCGCATGCTACAAACGATGATGCACAGATCCAGGCTTCTGGGCTCAGTGCCTC 585
 DB 518 LeuPheMetAsnSerTyrrLys-----MetLysLeuSerIleIleLeuGly---ValIle 534
 QY 586 CAGCTGACCTATCAGCTATGTGAGCTGATCTCTGCAGAGGTTCCCTCGGTGAGAGTT 645
 DB 535 HisMetThrPheCysLeuPheLeuSerLeuSerAsnTyrrArgPhePheLysArgLysLeu 554
 QY 646 GTGCTAATGGTATTTTCCCTGGTATCTGTACCTATGGGGCCACCTTTGCAATATGTTG 705
 DB 555 AspileTyrrAlaValPheValProSerLeuIlePheLeuGluAlaIlePheGlyTyrrLeu 574
 QY 706 GCTATCAGATCAAGTACGAT-----GACTACAGAGATTCCG---CTTGGGCCACTA 753
 DB 575 ValIleThrIleValTyrrLysTyrrCysIleAspTrpLysAlaLysAspLeuGlnProPro 594
 QY 754 GAAGTCTCTGCATCACCATCTGGCGGACATTTGGAGATCACTTCCCGCTCTGATCTG 813
 DB 595 SerLeuLeu-----AsnMetLeuLeuMet 603
 QY 814 GTGCTCTTCTCAGGCACCTTTGAAATTGAAGCTGTGCC-----TTCTAGTGTCTC 864
 DB 604 PheLeuSerProGlyThrLeuGluAspGlnLeuTyrrProGlyGlnLysTyrrLeuGlnVal 623
 QY 865 AACTTCTGTATC-----ATCCTCTTTGAGCCCTGGATT----- 897
 DB 624 GlyLeuValIleAlaAlaLeuIleCysValProTrpLeuLeuIleValLysProPheVal 643
 QY 898 -----AAGTCTCGAAGTGTGGCCATGCC 927
 DB 644 LeuTrpArgArgHisSerAsnGluAlaAsnLysTyrrGlnSerLeuAsnSerAspLeuPro 663
 QY 928 AAT-----AACATTGAAACAACTTCAGC--- 951

DB 664 AsnValAspGluAlaAspAlaLeuMetAlaValAspSerGlnGluLysGlnAlaGluPro 683
 QY 952 ---CGGCTCGGCACTCTGCTGCTC----- 972
 DB 684 PheGluLeuGlyGluValValIleHisGlnValIleHisThrIleGluPheCysLeuGly 703
 QY 973 CTGATTTCACTCACCATCTCTATGCTGCATCACTTCTCTTGTGTCAGCTGTTGCAG 1032
 DB 704 CysValSerHisThrAlaSerTyrrLeuArgLeu-----TrpAla----- 716
 QY 1033 TTGAGCTTTGGCACAAGATCTCTGTCGAAAGGCGACAGCACTGGGACATATGGCCCTG 1092
 DB 717 LeuSerLeuAlaHisAsnGlnLeu-----SerSerValLeuTrp---AsnMetThrLeu 733
 QY 1093 CACTATAGTGTGAGG-----TTGGTAGAAGATGTGATCATGCTGTGTTGTTTAAAG 1143
 DB 734 AlaAsnGlyPheArgMetThrGlyIleValGlySerIlePheValIleLeuPheGly 753
 QY 1144 TTCTTTGGAGTGAAGTGTGTACTGATTAATCTGTCATCTCTGATTCCTGTCAGCTCAT 1203
 DB 754 PheTrpPheIleAlaThr-----CysValValLeuValAlaMetGluGlyThr 769
 QY 1204 ATGCTTATCTGATTTCCATT 1224
 DB 770 SerAlaMetLeuHisSerLeu 776

RESULT 13
 MCSR_RAT STANDARD; PRT; 325 AA.
 ID MCSR_RAT
 AC P35345;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Melanocortin-5 receptor (MCS-R).
 GN MCSR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Striatum;
 RX MEDLINE=94234987; PubMed=8179577;
 RA Griffon N., Mignon V., Facchinetti P., Diaz J., Schwartz J.C.,
 RA Sokoloff P.;
 RT "Molecular cloning and characterization of the rat fifth melanocortin
 receptor.";
 RL Biochem. Biophys. Res. Commun. 200:1007-1014 (1994).
 CC -!- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA AND GAMMA) AND ACTH. THE
 CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
 CC ADENYLATE CYCLASE. THIS RECEPTOR IS A POSSIBLE MEDIATOR OF THE
 CC IMMUNOMODULATION PROPERTIES OF MELANOCORTINS.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: VERY LOW EXPRESSION LEVELS IS DETECTED IN
 CC BRAIN, WHILE HIGH LEVELS ARE FOUND IN ADRENALS, STOMACH, LUNG AND
 CC SPLEEN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L27081; AAA41577.1; -.
 DR PIR; JC2193; JC2193.
 DR InterPro; IPR000276; GPCR_Rhodpsn.
 DR Pfam; PF00001; 7tm1; 1.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
 DR PROSITE; PS02362; G-PROTEIN RECEPTOR FL2; 1.
 DR PROSITE; PS02362; G-PROTEIN RECEPTOR FL2; 1.

Qy	931	AACATTGAGAAAACTTCACCGGGTGCAGCACTCTGGTGCTCTGATTTTCAGTCACCATC	990
Db	227	SerValArgInAlaSerMetIysGlyala-----IlThrLeuThrMet	242
Qy	991	CTCATGTCTGGCATCAACTCTCTTCTGCTGGTCAGCTTTGCAGTTGAGGTTGGCAGACAGA	1050
Db	243	Leu---LeuGlYilePheIleValCysTrpSerProPhePheLeuHisLeuIle-----	259
Qy	1051	GATCTCGTCGACAAGGGCAGAACTGGGGACATATGGCGCTGCACATATAGTGTGAGGTG	1110
Db	260	-----LeuMetIleSerCys	264
Qy	1111	GTAGAGAATGTGATCATGCTCTTGGTTTTTAAGTTCTTTTGAGTGAAAAGTGTACTGAAT	1170
Db	265	ProGlnAsnValTyrcysAlaCysPheMetSerTyrrPheAsnMetTyrrLeuIleLeuIle	284
Qy	1171	TACTGTCAATCTCTTGATTCGCCCTTGCAGCTCATATTATGCTTATCTGATTTCCATTGGCTTC	1230
Db	285	MetCysAsnSerValIleasp-----	291
Qy	1231	ATGCTCCTTTTCTTCAGTATCTTGCATCTGCATCTGGCTGCACCTCTTACCCTAATAAGTAGTA	1290
Db	292	-----ProLeuIleTyrrAlaLeuArgSerGlnGluMetArgArgThr---	305
Qy	1291	GACTACCTCCATTGTCTGCTGTGCACGAGCACCTCGGACC	1332
Db	306	---PheLysGluIleIleCysHisGlyPheArgArgThr	318
RESULT 14			
ID	NU2M	CHLRE	
ID	NU2M	CHLRE	STANDARD; PRT; 382 AA.
AC	P08740;	P11659;	
DT	01-AUG-1988	(Rel. 08, Created)	
DT	01-AUG-1988	(Rel. 08, Last sequence update)	
DT	15-JUL-1999	(Rel. 38, Last annotation update)	
DE	NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).		
DE	ND2 OR NAD2.		
OS	Chlamydomonas reinhardtii.		
OG	Mitochondrion.		
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
OC	Chlamydomonadales; Chlamydomonades.		
OX	NCBI_TaxID=3055;		
[1]			
RP	SEQUENCE FROM N.A.		
RX	STRAIN=CW15-2;		
RX	MEDLINE=87016388; PubMed=3020517;		
RA	Boer P.H., Gray M.W.;		
RA	"Nucleotide sequence of a protein coding region in Chlamydomonas		
RL	reinhardtii mitochondrial DNA.";		
RL	Nucleic Acids Res. 14:7506-7507(1986).		
[2]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cw15;		
RA	Gray M.W.;		
RA	Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.		
[3]			
RP	SEQUENCE OF 305-382 FROM N.A.		
RC	STRAIN=cw15;		
RX	MEDLINE=89083574; PubMed=3205746;		
RA	Ma D.-P., Yang Y.-W., Hasnain S.;		
RL	"Nucleotide sequence of Chlamydomonas reinhardtii mitochondrial genes		
RT	coding for subunit 6 of NADH dehydrogenase and trnaTrp.";		
RL	Nucleic Acids Res. 16:11373-11373(1988).		
[4]			
RP	SEQUENCE OF 366-382 FROM N.A.		
RC	STRAIN=cw15;		
RX	MEDLINE=89282417; PubMed=2734113;		
RA	Boer P.H., Gray M.W.;		
RT	"Nucleotide sequence of a region encoding subunit 6 of NADH		
RT	dehydrogenase (ND6) and trna(Trp) in Chlamydomonas reinhardtii		
RL	mitochondrial DNA.";		
RL	Nucleic Acids Res. 17:3993-3993(1989).		

Db 40 LysPheAsnPhetyrPheAsnSerIleLeuPheSerPheIleLeuPheLeu-----Val 57

```
QY 184 GCATCTGCTTTGTATCATGGTTAGAAATCTCTCAAGAAAGTACTGAAACTTACTGGATGACA 243
Db 58 ThrPheSerValLeuValPheSerThrTyTyTyTyLeuAsnSerGlu----- 72
QY 244 TACACCTTTCTTTCTTTATGTTTTCATCCATTATGGTCCAGTTGACCCCTCAATTTTGTGTC 303
Db 73 LeuAsnPheAsnTyTyTyTyPhe-----ValLeuLeuLeuPheVal 86
QY 304 CACAGAGATCTAGCCAAAGATAAACCCTATCATATTATTATGCTTAATCTCTCTGGGA 363
Db 87 GlySerMetPheSerLeuAsnPheSerAsnSerIlePheThrMetLeuLeu-Leu-SerTrpAs 106
QY 364 CTTGTTATCAGA----- 380
Db 106 pLeuLeuGlyIleSerSerPhePheLeuValLeuPheTyTyAsnAsnTrpAspSerCysSe 126
QY 381 GGAGGCCATGATTAAAGTACCTCACACTGTGGAAGAAAGAGGAGGAGGAGGCCCTATGT 440
Db 126 rGlyAlaMetAsnThrAlaLeuThr-----AsnArgLeuGlyAspTyTyPh 141
QY 441 CAGCCTCACCCGAAAGAGATGCTTAATAGATGGCGAGGAGTGTCTGATAGAAATGGGAGGT 500
Db 141 eMetPheValPhePheGlyLeuSerValPheSerGlyTyTyTyPheLeuSerPheSerMe 161
QY 501 GGCCCACTCCATCCGACCTGGGTATGCACCCGCAATGCCACAAACGTATGTCCACAGAT 560
Db 161 tPheSerSerTyMetSerLeuLeuLeuLeuLeuLeuAlaPheThrLySerAlaGlnPh 181
QY 561 CCAAGCCTTCTGGGCTCAGTGGCCGACGCTGACCTATCAGCTCTATGTGAGCCGTGATCTC 620
Db 181 e---ProPheSerSerTrpLeuProLys-----AlaMetSe 192
QY 621 TGCAGAGGTTCCCTGGGTAGATGTGTAAATGGTATTTTCCCTGTATCTGTACACCTA 680
Db 192 rAlaProThrPro-----ValSerSerLeuValHisSerSerThrLeuValThrAl 209
QY 681 TGGGGCCACCCCTTTGC-----AATATGTTGGCTATCCAGATCAAGTACGATGACTA 731
Db 209 aGlyLeuLeuLeuLeuMetAsnPheAsnLeuValMetGlnLysAspPheIleSerPh 229
QY 732 CAAGATTCGCCCTTGGG----- 747
Db 229 eValLeuIleIleGlyLeuPheThrMetPhePheSerSerLeuAlaSerLeuValGluGl 249
QY 748 -----CCACTAGAAGTCTCTC-----TGCATCAC 770
Db 249 uAspLeuLysLysValValAlaLeuSerThrLeuSerGlnMetGlyPheSerMetValTh 269
QY 771 CATCTGGCGGACATTTGGAGATCACTTCC-----CGCCTCTCTGATTCTGTGCTCTTCTC 824
Db 269 rLeuGlyLeuGlyLeuSerPheIleSerPheIleHisLeuValSerHisAlaLeuPheLy 289
QY 825 AGCCACTTTGAATTTGAAGCT----- 846
Db 289 sSerCysLeuPheMetGlnValGlyTyTyIleIleHisCysSerPheGlyGlnGlnAspGl 309
QY 847 -----GTGCCC-----TTCTAGTGTCTCAACTCTCTGATCAT 878
Db 309 yArgAsnTyTySerAsnAsnGlyAsnLeuProAsnPheIleGlnLeuGlnMetLeuValTh 329
QY 879 CTTCTTTGAGCCCTGGATTAAAGTTCTGGAGAAGTGGTGGCCAGATGCCCAATACATT-- 936
Db 329 rLeuPheCysLeuCysGlyLeuIlePheSerSerGlyAlaValSerLysAspPheIleLe 349
QY 937 -----GAGAAACACTTCACCGGGTCGGCACTCTGGTGGTCTCTGATTTTCAGT 983
Db 349 uGluLeuPhePheSerAsnAsnTyTyMetPhePheSerSerLeuMetPhePheValSerVa 369
QY 984 CACCATCTCTATGCTGGCATCAACTTCTCTGTGTGTGTCAGCTTTGCGAGTTGAGTTGGC 1043
Db 369 lPheLeuThrPhe---GlyTyTySerPheArgLeuTrpLysSerPhePheLeuSerPhe-- 387
QY 1044 AGACAGAGATCTCGTCGACAAAGGGCAGAACTGGGGACATATGGGCTGCACTATAGTGT 1103
```

```
Db 388 -----AsnLysValMetAsn-----HisTySerSe 396
QY 1104 GAGTTGGTAGAGAAATGTGATCATGGTCTTGGTTTAAAGTTCTTTGGAGTGAAAGTGT 1163
Db 396 rThrValPheMetAsn---PheLeuSerLeuValLeuValIlePheSerIleSerPheLe 415
QY 1164 ACTGAATTACTGTCTATTCCTTGGATTGGCTTGCAGCTCATTTATTGCTTATCTGATTCCAT 1223
Db 415 uTrpTrpMetAsnPheAsnLeuLeuAsnIleProSerLeuPheLeuTyTyValAspPhePh 435
QY 1224 TGGC-----TTCATGCTCCTTTTCTTCCAGTACTTGCATCCATCCATGCGCTC 1268
Db 435 eGlyProLeuValPheLeuPheMetMetIlePheLeuSerPheLeu---IleLeuLysMe 454
QY 1269 ACTCTTCACCCAT-----AATGTAGTAGACTACCTC 1299
Db 454 tLeuPheLysGluLeuMetTyTyLysPheLeuValAspTyTyLeu 468
```

Search completed: April 1, 2003, 08:54:20
Job time : 36.5 secs